Supplemental Data

Table S1, related to Figure 1. Clinical and biological features of the MCL patients included in the present study.

Sample	Age	Sex	*IGHV* Identity (%)	SOX11	Sample Type	Empiric sample purity (%)	Method (Sample purity)	Sample purity based on DNA methylation (%)	Sampling time	Morphology	Nodal	Mutations	Gene expression data	Treatment	Status	Series	Epigenetic classification
M001	63	1	98.7	1	LN	90.0%	Visual inspection-IHC	94.9%	At diagnosis	classic/blastic	1		1	1	1	BCN	C1 MCL
M003	75	1	93.72	0	PB	96.0%	Flow cytometry	94.6%	At diagnosis	pleomorphic	0	*TLR2, TP53*	1	1	1	BCN	C2 MCL
M004	81	1	97.22	0	PB	99.4%	Flow cytometry	99.0%	Pre-treatment	classic	0	*TP53*		1	0	BCN	C2 MCL
M009	78	0	92.36	0	PB	98.0%	Flow cytometry	90.9%	Post-treatment	small cell	0	*BIRC3*	1	1	1	BCN	C2 MCL
M012	59	0	96.18	1	LN	82.0%	Flow cytometry	76.6%	At diagnosis	classic	1	*MEF2B*	1	1	0	BCN	C1 MCL
M015	70	1	96.53	0	PB	85.2%	Flow cytometry	80.4%	Untreated	small cell	0		1	0	0	BCN	C2 MCL
M016	50	0	92.04	0	PB	97.2%	Flow cytometry	97.9%	Pre-treatment	small cell	0		1	0	0	BCN	C2 MCL
M021	73	1	94.44	0	PB	99.0%	Flow cytometry	98.2%	At diagnosis	small cell	0	*TLR2*	1	0	0	BCN	C2 MCL
M026	63	1	95.09	1	PB	99.5%	Flow cytometry	99.1%	At diagnosis	classic	1	*TP53*	1	1	1	BCN	C1 MCL
M027	63	1	96.88	0	PB	98.9%	Flow cytometry	97.7%	Untreated	small cell	0	*TP53*	1	0	0	BCN	C2 MCL
M029	68	1	98.25	1	PB	99.3%	Flow cytometry	97.6%	At diagnosis	blastoid/pleomorphic	0	*NOTCH2, TP53*	1	1	1	BCN	C1 MCL
M030	89	0	98.26	1	PB	99.7%	Flow cytometry	98.7%	At diagnosis	classic	1		1	1	1	BCN	C1 MCL
M035	54	0	100	1	LN			79.7%	At diagnosis	classic	1			1	1	BCN	C1 MCL
M039	82	0	87.68	0	PB	95.0%	Flow cytometry	90.2%	At diagnosis	small cell	0	*TP53*	1	1	1	BCN	C2 MCL
M047	71	1	98.63	1	LN			85.1%	At diagnosis	blastoid	1			1	0	BCN	C1 MCL
M054	71	1	97.31	1	LN			78.0%	At diagnosis	pleomorphic	1			1	1	BCN	C1 MCL
M059	68	1	99.55	1	LN			88.9%	Post-treatment	classic	1			1	1	BCN	C1 MCL
M071	51	0	99.12	1	LN			63.8%	At diagnosis	classic	1			1	1	BCN	C1 MCL
M075	58	1	99.09	0	mucosa			80.5%		classic						BCN	C1 MCL
M076	66	1	93.68	0	PB	68.0%	Flow cytometry	69.9%	At diagnosis	small cell	0		1	0	0	BCN	C2 MCL
M078	68	0	99.55	1	LN			87.2%	Post-treatment	pleomorphic	1	*TP53*		1	1	BCN	C1 MCL
M088	66	1	100	1	spleen			85.9%	Post-treatment	classic	0	*TP53, WHSC1*	1	1	1	BCN	C1 MCL
M106	67	1	100	1	LN			75.4%	At diagnosis	classic	1			1	1	BCN	C1 MCL
M108	63	1	100	1	tissue			88.5%	At diagnosis	classic	1			1	0	BCN	C1 MCL
M114	61	0	100	1	spleen			75.8%			1		1	1	1	BCN	C1 MCL
M122	70	1	99.55	1	LN			77.9%	At diagnosis	classic	1	*MEF2B*	1	1	0	BCN	C1 MCL
M163	43	1	92.04	0	LN			67.0%	At diagnosis	classic	1			0	0	BCN	C2 MCL
M195	69	1	93.75	0	PB	93.0%	Flow cytometry	95.2%	At diagnosis		1	*TP53*		1	1	BCN	C2 MCL
M197	64	1	96.53	0	PB	86.0%	Flow cytometry	80.5%	At diagnosis	blastoid	0	*TP53*		1	1	BCN	C2 MCL
M198	49	0	100	1	PB	85.0%	Flow cytometry	72.8%	At diagnosis	classic	1	*WHSC1*	1	1	1	BCN	C1 MCL
M199	65	1	94.44	1	PB	89.0%	Flow cytometry	81.7%	Pre-treatment	classic	1	*TP53, WHSC1*	1	1	1	BCN	C1 MCL
M200	70	1	98.61	1	PB	90.0%	Flow cytometry	81.3%	At diagnosis	blastoid	1	*NOTCH2, TP53, WHSC1*	1	1	1	BCN	C1 MCL
M201	85	1	97.19	0	PB	90.0%	Flow cytometry	92.6%	Pre-treatment	classic	0		1	1	1	BCN	C2 MCL
M202	51	1	99.65	1	PB	72.0%	Flow cytometry	67.7%	At diagnosis	classic	1			1	0	BCN	C1 MCL
M203	58	1	99.33	1	PB	94.0%	Flow cytometry	94.2%	At diagnosis	classic	0		1	1	0	BCN	C1 MCL
M205	64	1	100	1	PB	80.0%	Flow cytometry	69.5%	At diagnosis	classic	1			1	0	BCN	C1 MCL
M206	64	0	100	1	PB	92.0%	Flow cytometry	97.7%	At diagnosis	classic	1			1	0	BCN	C1 MCL
M207	69	1	98.95	1	PB	68.0%	Flow cytometry	63.0%	At diagnosis	classic	1			1	0	BCN	C1 MCL

Sample	Age	Sex	*IGHV* Identity (%)	SOX11	Sample Type	Empiric sample purity (%)	Method (Sample purity)	Sample purity based on DNA methylation (%)	Sampling time	Morphology	Nodal	Mutations	Gene expression data	Treatment	Status	Series	Epigenetic classification
M208	63	1	98.96	1	PB	62.0%	Flow cytometry	63.9%	At diagnosis	classic	1			1	0	BCN	C1 MCL
M209	42	1	98.96	1	PB	56.0%	Flow cytometry	53.6%	At diagnosis	blastoid	1	*WHSC1*		1	1	BCN	C1 MCL
M210		1		1	PB	81.0%	Flow cytometry	80.3%	At diagnosis	classic	1		1		0	BCN	C1 MCL
M211	56	1	97.57	0	PB	86.0%	Flow cytometry	84.2%	Pre-treatment	classic	0			1	0	BCN	C2 MCL
M212	60	1	100	1	PB	84.0%	Flow cytometry	79.9%	At diagnosis		1	*BIRC3*	1	1	0	BCN	C1 MCL
M213	70	1	100	0	PB	68.0%	Flow cytometry	51.1%	Pre-treatment	classic	0	*TP53, WHSC1*		0	0	BCN	C1 MCL
M214	57	1	97.57	1	PB	84.0%	Flow cytometry	77.1%	At diagnosis	classic	0	*WHSC1*	1	1	0	BCN	C1 MCL
M215	54	1	100	1	PB	60.0%	Flow cytometry	39.5%	At diagnosis	classic	1			1	0	BCN	C1 MCL
M218	69	1	94.74	0	PB	100.0%	Flow cytometry	100.8%	At diagnosis		0		1	0	0	BCN	C2 MCL
M220	76	1	97.57	0	BM			71.1%	At diagnosis						1	BCN	C2 MCL
M221	54	1	100	1	BM	89.5%	Flow cytometry	87.8%	At diagnosis			*BIRC3*			1	BCN	C1 MCL
M224	72	1	100	0				92.7%	At diagnosis			*MEF2B*			1	BCN	C1 MCL
M225	71	1	100	1	BM	90.0%	Flow cytometry	68.3%	At diagnosis					1	1	BCN	C1 MCL
M227	61	1	97.76					43.4%	At diagnosis						0	BCN	C1 MCL
M228	68	1	91.67	0	BM			52.6%	At diagnosis						0	BCN	C2 MCL
V17724		1	98.7	1	LN	90.0%	Visual inspection-IHC	67.4%			1					Kiel	C1 MCL
V17782		0	100	1	LN	80.0%	Visual inspection-IHC	76.2%	At diagnosis		1					Kiel	C1 MCL
V17824	75	1	99.1	1	LN	90.0%	Visual inspection-IHC	62.7%	At diagnosis		1			1	1	Kiel	C1 MCL
V17832		1	97.8		LN	80.0%	Visual inspection-IHC	72.9%	At diagnosis		1					Kiel	C1 MCL
V17985		1	99.6	1	LN	90.0%	Visual inspection-IHC	80.1%	At diagnosis		1					Kiel	C1 MCL
V18094	62	1	99.1	1	LN	80.0%	Visual inspection-IHC	81.1%	At diagnosis		1			0	0	Kiel	C1 MCL
V18131		1	92.3	0	LN	70.0%	Visual inspection-IHC	59.3%			1					Kiel	C2 MCL
V18213		1	99.6		LN	70.0%	Visual inspection-IHC	68.6%	At diagnosis		1					Kiel	C1 MCL
V18258	69	1	97.5		LN	80.0%	Visual inspection-IHC	72.0%	At diagnosis		1			1	0	Kiel	C1 MCL
V18348		1	100	1	LN	90.0%	Visual inspection-IHC	73.8%			1					Kiel	C1 MCL
V18485		1	99.1	1	LN	90.0%	Visual inspection-IHC	80.5%			1					Kiel	C1 MCL
V18498	76	0	98.3	1	LN	90.0%	Visual inspection-IHC	77.8%	Pre-treatment		1			1	1	Kiel	C1 MCL
V18505		1	97.8	1	LN	90.0%	Visual inspection-IHC	79.5%			1					Kiel	C1 MCL
V18562	55	1	100	1	LN	90.0%	Visual inspection-IHC	82.8%	At diagnosis		1			1	1	Kiel	C1 MCL
V18573	56	0	99.1	1	LN	90.0%	Visual inspection-IHC	77.3%	At diagnosis		1			0	0	Kiel	C1 MCL
V18587	73	1	98.96		LN	90.0%	Visual inspection-IHC	70.8%	At diagnosis		1			1	0	Kiel	C1 MCL
V18630	84	1	100		LN	90.0%	Visual inspection-IHC	77.9%	At diagnosis		1			1	0	Kiel	C1 MCL
V18633		1	100	1	LN	70.0%	Visual inspection-IHC	76.2%	At diagnosis		1					Kiel	C1 MCL
V18663		1	100	1	LN	80.0%	Visual inspection-IHC	80.3%	Pre-treatment		1					Kiel	C1 MCL
V18711	34	1	100	1	LN	90.0%	Visual inspection-IHC	83.1%	Pre-treatment		1					Kiel	C2 MCL
V18719		1	94.3	1	LN	80.0%	Visual inspection-IHC	78.5%			1					Kiel	C1 MCL
V18726		1	97.8	0	LN	80.0%	Visual inspection-IHC	75.8%			1					Kiel	C2 MCL
V18737	60	1	100	1	LN	80.0%	Visual inspection-IHC	76.2%	At diagnosis		1				0	Kiel	C1 MCL
V18750		1	100	1	LN	90.0%	Visual inspection-IHC	67.6%			1					Kiel	C1 MCL
V18917	85	1	98.7	1	LN	80.0%	Visual inspection-IHC	74.4%	At diagnosis		1			0	1	Kiel	C1 MCL
V19080	57	0	99.1		LN	80.0%	Visual inspection-IHC	68.8%	At diagnosis		1			0	0	Kiel	C1 MCL
V19164		1	100	1	LN	90.0%	Visual inspection-IHC	88.9%	At diagnosis		1					Kiel	C1 MCL
V19165		1	99.6	0	LN	80.0%	Visual inspection-IHC	67.5%			1					Kiel	C1 MCL
V19203		1	99.1	1	LN	90.0%	Visual inspection-IHC	80.1%	Pre-treatment		1					Kiel	C1 MCL

Sex: 0-female, 1-male; SOX11: 0-not expressed, 1-expressed; Nodal: 0-non-nodal, 1-nodal; Treatment: 0-no treatment, 1-treated; Status: 0-alive, 1-dead, 2-unknown. PB: peripheral blood; LN: lymph node; BM: bone marrow

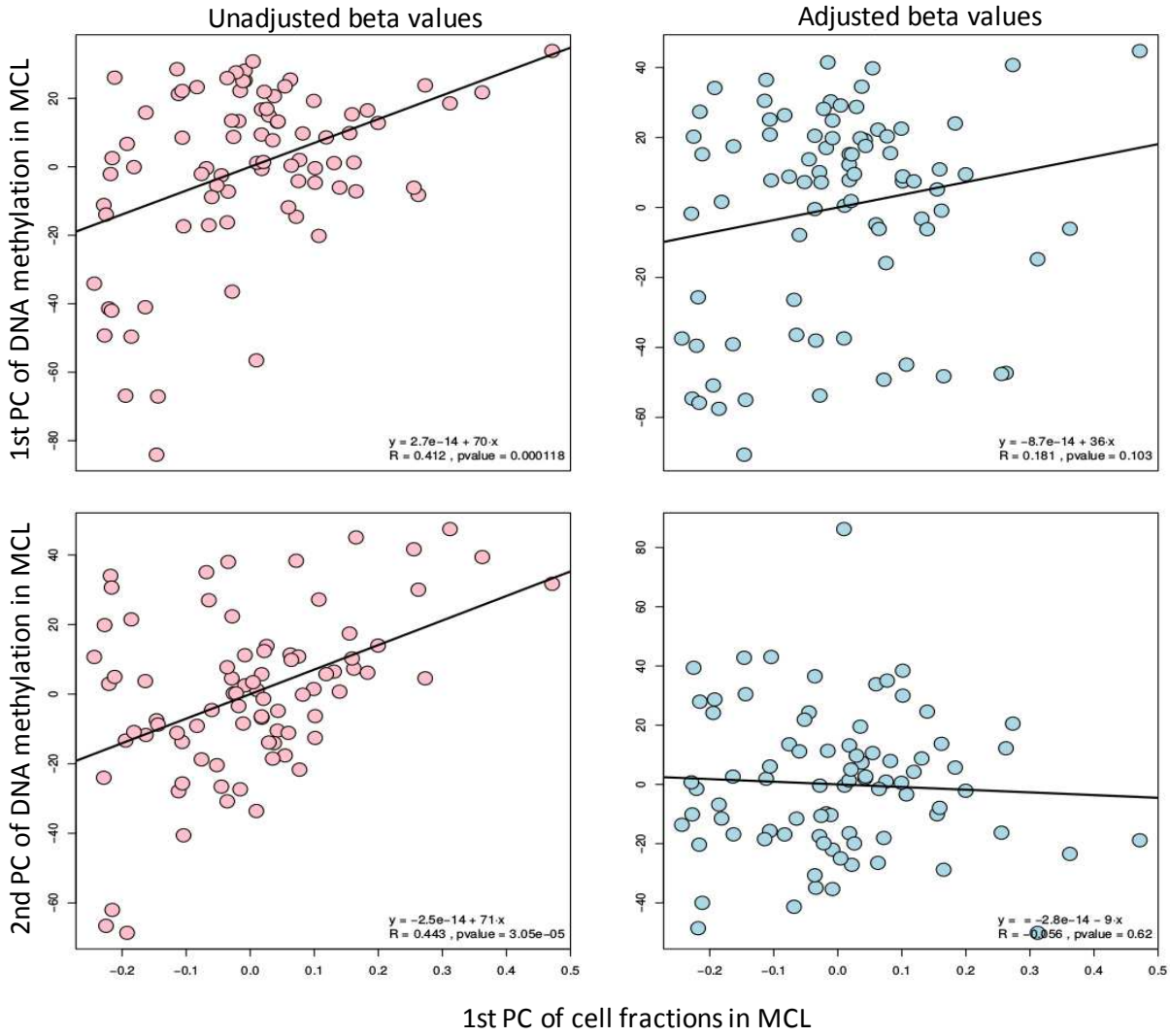


Figure S1, related to Figure 1. Effect of cell composition in DNA methylation variability.

First and second Principal Component of DNA methylation in MCL samples against first Principal Component of the proportion of all 6 hematopoietic cell types in MCL before (left) and after (right) correcting DNA methylation estimates according to MCL purities. PC, principal component.

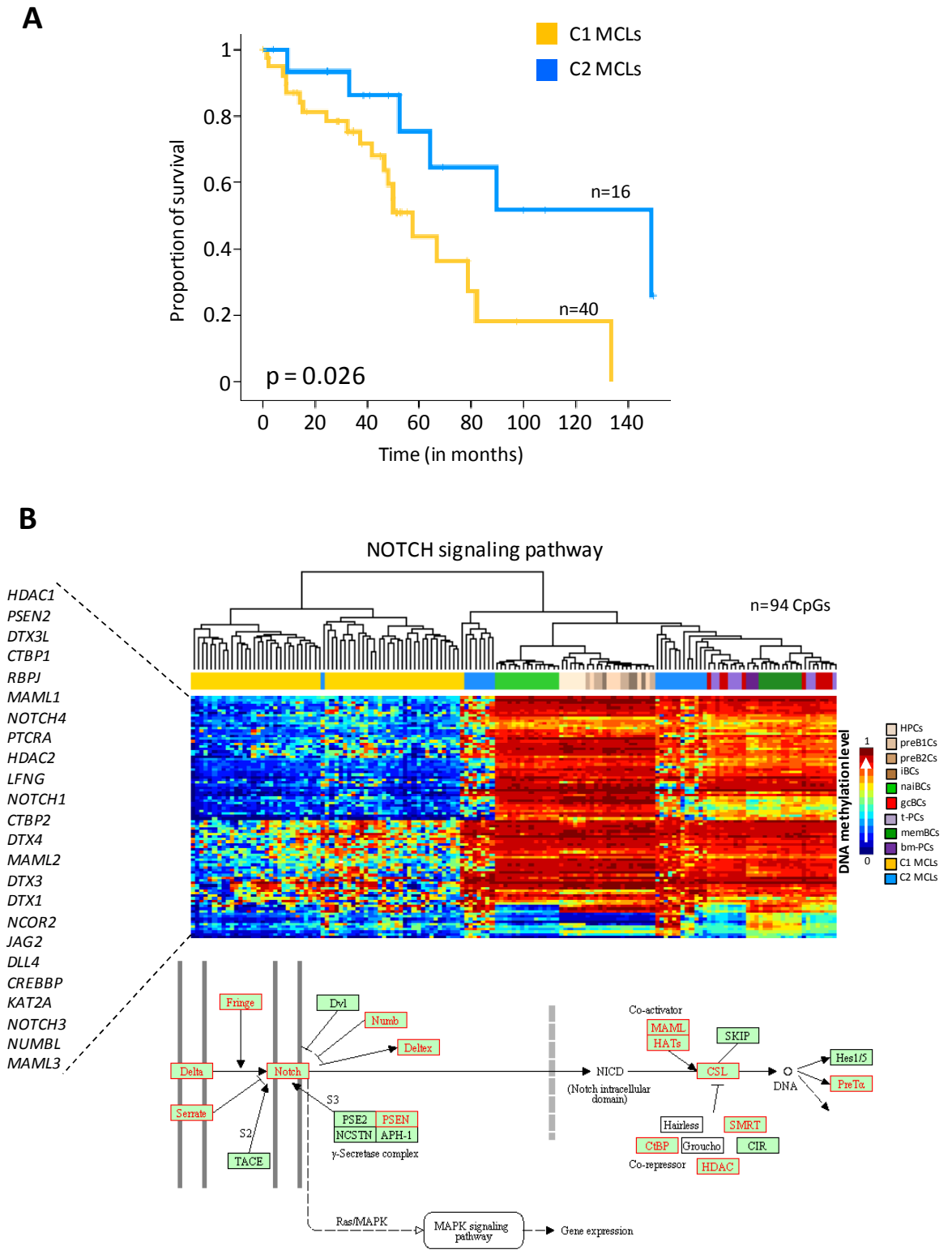


Figure S2, related to Figure 2. Biological data between the two MCL subgroups.
(A) Kaplan-Meier plot showing that, as compared with C2, C1 MCLs globally show a worse overall survival. Only C1 and C2 MCL cases with available full clinical reports were used for this analysis (n=56). **(B)** Heatmap representation of the differentially methylated CpGs between MCL C1 and C2 belonging to genes in the NOTCH signaling pathway (upper panel) and the exact location of these genes within the NOTCH pathway (lower panel). Genes containing differentially methylated CpGs are presented in red in the graphical representation whereas genes with stable DNA methylation are represented in black.

Table S2, related to Figure 2. Pathways altered in MCL in regions hypomethylated between cluster 1 MCLs and cluster 2 MCLs.

Pathway_Name	Number of genes found	Number of genes in pathway	Percentage	pvalue	pvalueAdjusted
Axon guidance	72	176	0.41	5.63E-13	8.53E-11
Notch signaling pathway	24	48	0.5	1.73E-07	1.74E-05
Neurotrophin signaling pathway	45	120	0.38	2.29E-07	1.74E-05
Phospholipase D signaling pathway	50	144	0.35	8.40E-07	5.09E-05
Glycerolipid metabolism	26	59	0.44	1.54E-06	7.79E-05
Thyroid hormone signaling pathway	41	118	0.35	7.13E-06	0.000308795
Proteoglycans in cancer	62	205	0.3	9.10E-06	0.000344638
Estrogen signaling pathway	35	99	0.35	1.93E-05	0.000585795
Chronic myeloid leukemia	28	73	0.38	1.91E-05	0.000585795
Fc gamma R-mediated phagocytosis	33	93	0.35	2.91E-05	0.00080148
Glycosaminoglycan biosynthesis - chondroitin sulfate / dermatan sulfate	11	20	0.55	4.90E-05	0.00092822
Endocytosis	72	260	0.28	4.70E-05	0.00092822
Inflammatory mediator regulation of TRP channels	34	98	0.35	3.90E-05	0.00092822
Pathways in cancer	101	397	0.25	4.87E-05	0.00092822
Glioma	25	65	0.38	4.50E-05	0.00092822
Aldosterone synthesis and secretion	29	81	0.36	6.54E-05	0.001165383
Glycerophospholipid metabolism	32	95	0.34	0.00012041	0.002026921
ErbB signaling pathway	30	88	0.34	0.00014486	0.002310122
Wnt signaling pathway	43	143	0.3	0.00022235	0.003368661
Vascular smooth muscle contraction	37	120	0.31	0.00032445	0.004096136
GnRH signaling pathway	30	91	0.33	0.00028717	0.004096136
Prolactin signaling pathway	25	72	0.35	0.00031912	0.004096136
Oxytocin signaling pathway	46	158	0.29	0.00031744	0.004096136
Ras signaling pathway	61	228	0.27	0.00048343	0.005859209
Cholinergic synapse	34	111	0.31	0.00061064	0.007116258
Circadian rhythm	13	31	0.42	0.00063609	0.007138295
Acute myeloid leukemia	20	57	0.35	0.00088431	0.009569493
EGFR tyrosine kinase inhibitor resistance	26	81	0.32	0.0010261	0.01072094
mTOR signaling pathway	42	149	0.28	0.00113183	0.011233692
Focal adhesion	54	203	0.27	0.00114932	0.011233692
Long-term potentiation	22	66	0.33	0.00122898	0.011636939
Chemokine signaling pathway	50	187	0.27	0.00151907	0.013947784
Dopaminergic synapse	37	130	0.28	0.0017224	0.015223583
Insulin signaling pathway	39	139	0.28	0.00179073	0.015223583
Non-small cell lung cancer	19	56	0.34	0.00180874	0.015223583
Phosphatidylinositol signaling system	29	98	0.3	0.00247775	0.020290719
Dorso-ventral axis formation	11	28	0.39	0.00267473	0.021225755
Small cell lung cancer	26	86	0.3	0.00273203	0.021225755
Rap1 signaling pathway	54	211	0.26	0.00286951	0.021736521
Circadian entrainment	28	95	0.29	0.00305031	0.022542557
Morphine addiction	27	91	0.3	0.00316337	0.022821471
Melanogenesis	29	100	0.29	0.00345853	0.024370562
Platelet activation	34	122	0.28	0.00365204	0.025149273
Adrenergic signaling in cardiomyocytes	40	149	0.27	0.00380648	0.025630298
AGE-RAGE signaling pathway in diabetic complications	29	101	0.29	0.00406101	0.026749694
Fat digestion and absorption	14	41	0.34	0.00510983	0.032942107
ECM-receptor interaction	24	82	0.29	0.00591129	0.037315006
cGMP-PKG signaling pathway	43	167	0.26	0.0063645	0.038568895
Adherens junction	22	74	0.3	0.00633239	0.038568895
Insulin resistance	30	109	0.28	0.00701999	0.041706971
Histidine metabolism	9	24	0.38	0.00762491	0.04181364
Glycosphingolipid biosynthesis - globo series	6	14	0.43	0.00786593	0.04181364
Lysosome	33	123	0.27	0.00769395	0.04181364
Gap junction	25	88	0.28	0.00786279	0.04181364
T cell receptor signaling pathway	29	105	0.28	0.00742669	0.04181364
Choline metabolism in cancer	28	101	0.28	0.00784777	0.04181364
Cell adhesion molecules (CAMs)	38	146	0.26	0.00802903	0.041944757
Endocrine and other factor-regulated calcium reabsorption	15	47	0.32	0.00855316	0.043925546
cAMP signaling pathway	49	199	0.25	0.00943416	0.046861476
Transcriptional misregulation in cancer	45	180	0.25	0.0094136	0.046861476
Long-term depression	18	60	0.3	0.01021943	0.049943341

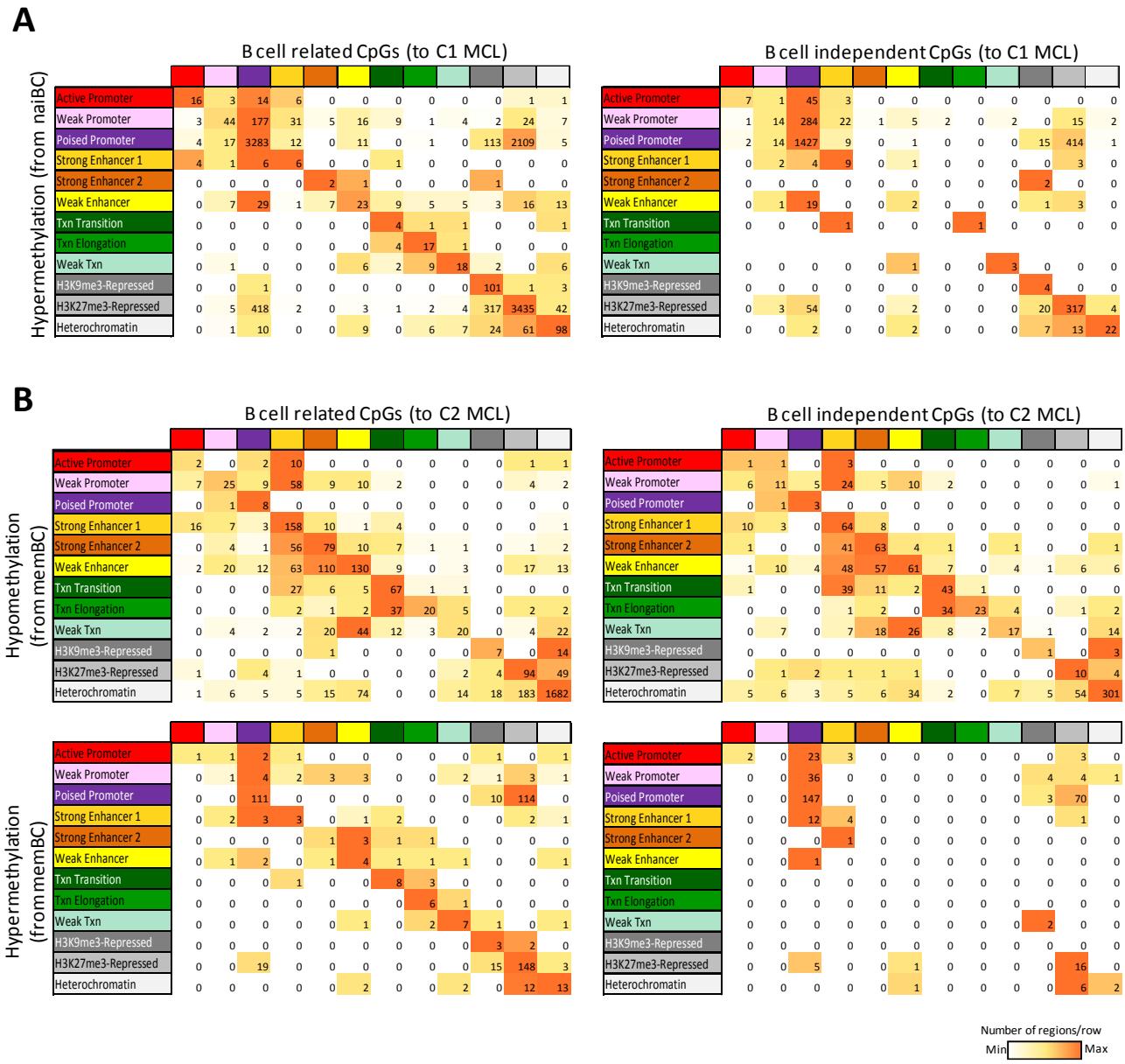


Figure S3, related to Figure 3. Transitions among chromatin states in relationship with DNA methylation changes.

(A) Transitions in chromatin states from naiBCs to C1 MCL at the B cell-related and B cell-independent hypermethylated CpGs. **(B)** Transitions in chromatin states from memBCs to C2 MCL at the B cell-related and B cell-independent hypomethylated and hypermethylated CpGs.

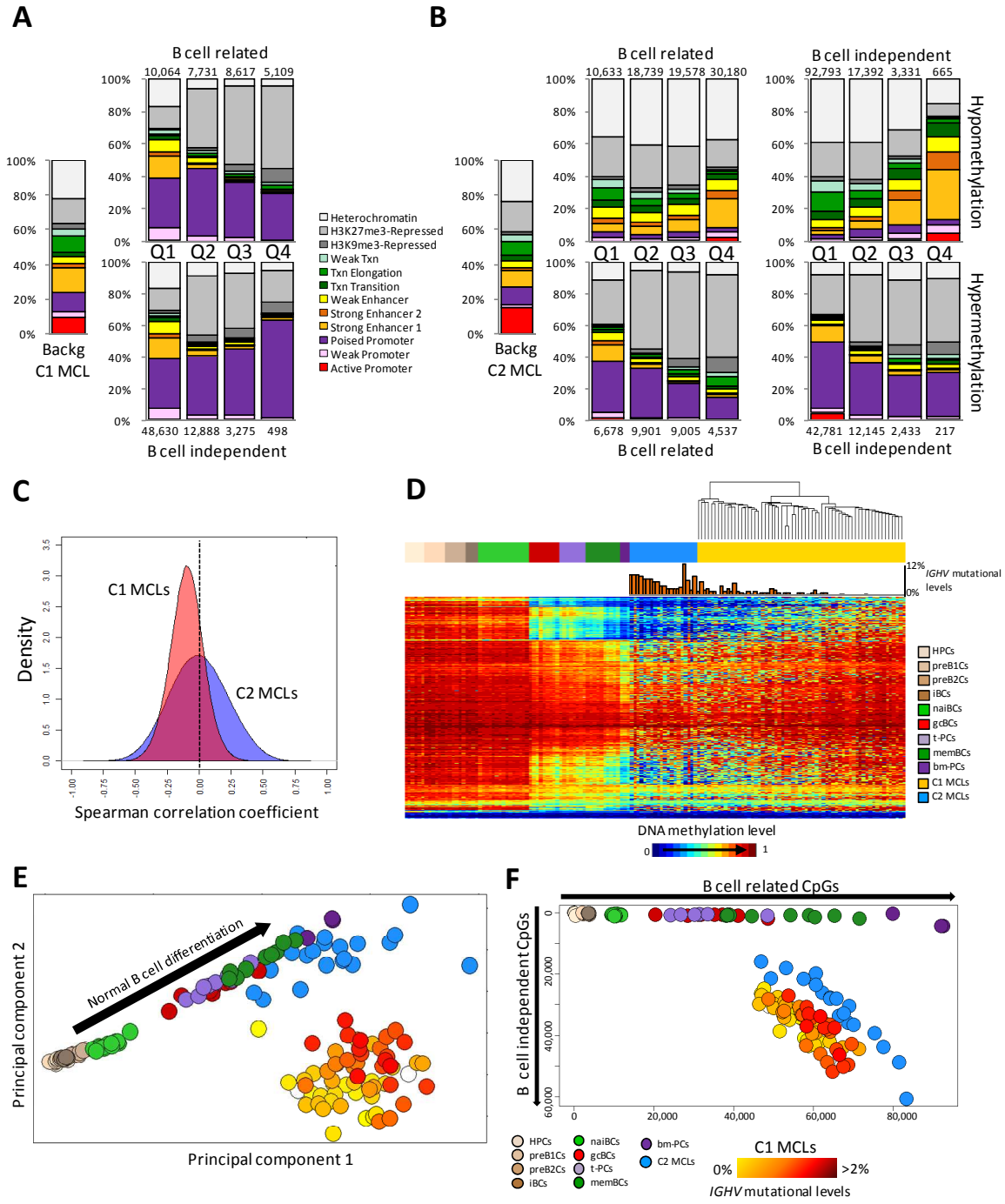
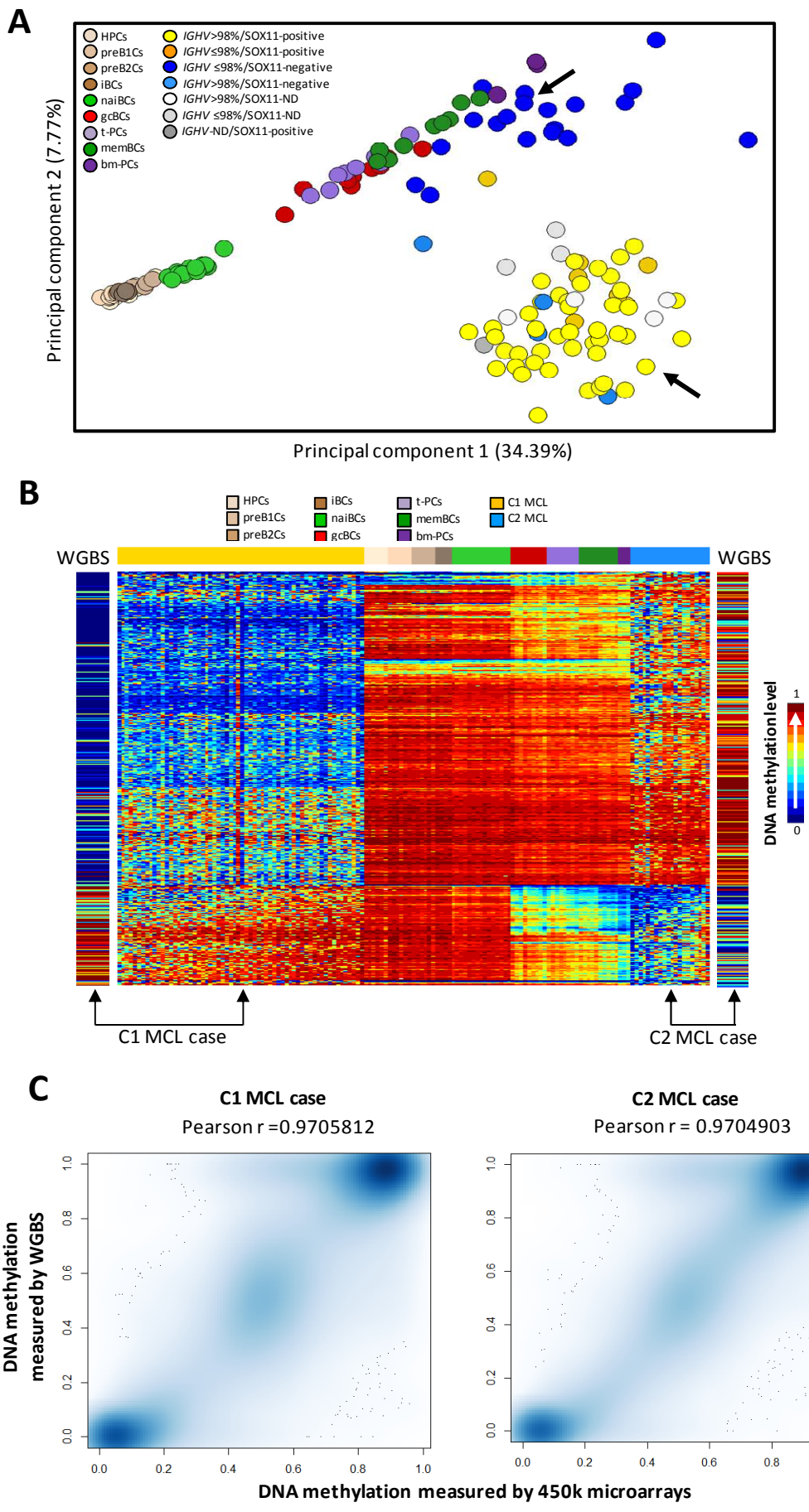


Figure S4, related to Figure 4. Chromatin states in recurrently altered CpGs and association between DNA methylation and *IGHV* mutational levels.

(A) Chromatin states, defined in a MCL primary case representative of C1 cases, of the hypermethylated CpGs between C1 MCLs and HPCs divided in quartiles based on their level of recurrence. Q1, recurrent in <25% of the cases; Q2, recurrent in 25-50% of the cases; Q3, recurrent in 50-75% of the cases; Q4, recurrent in >75% of the cases. **(B)** Chromatin states, defined in a MCL primary case representative of C2 cases, of the differentially methylated CpGs between C2 and HPCs divided in quartiles based on their level of recurrence. **(C)** Density plot showing the results of the correlation analysis between DNA methylation levels of all 450K CpGs and *IGHV* mutational levels, for both MCL C1 (red) and C2 (blue). C1 MCLs show a shift towards an inverse correlation whereas C2 MCLs show a distribution centered around 0 (i.e. no correlation). **(D)** Heatmap representation of the 6,245 CpGs

showing a significant negative correlation (correlation coefficient <-0.4 and FDR <0.05) with *IGHV* mutational levels in the C1 group. **(E-F)** New versions of the unsupervised PCA defined in Figure 2A **(E)** and scatter plot defined in Figure 4E **(F)** including a color code for C1 MCLs based on their somatic hypermutation level. These graphics indicate that already at the global and unsupervised level, C1 MCLs show different DNA methylation patterns depending of the level of *IGHV* somatic mutations and that these patterns follow the direction of the normal B cell differentiation.


A
HPCs
preB1Cs
preB2Cs
iBCs
naiBCs
gcBCs
t-PCs
memBCs
bm-PCs
IGHV>98%/SOX11-positive
IGHV≤98%/SOX11-positive
IGHV ≤98%/SOX11-negative
IGHV>98%/SOX11-negative
IGHV>98%/SOX11-ND
IGHV ≤98%/SOX11-ND
IGHV-ND/SOX11-positive
Principal component 2 (7.77%)
Principal component 1 (34.39%)
B
HPCs
preB1Cs
preB2Cs
iBCs
naiBCs
gcBCs
t-PCs
memBCs
bm-PCs
C1 MCL
C2 MCL
WGBS
WGBS
DNA methylation level
1
0
C1 MCL case
C2 MCL case
C
C1 MCL case
Pearson r =0.9705812
C2 MCL case
Pearson r = 0.9704903
DNA methylation measured by WGBS
DNA methylation measured by 450k microarrays
0.0
0.2
0.4
0.6
0.8
1.0

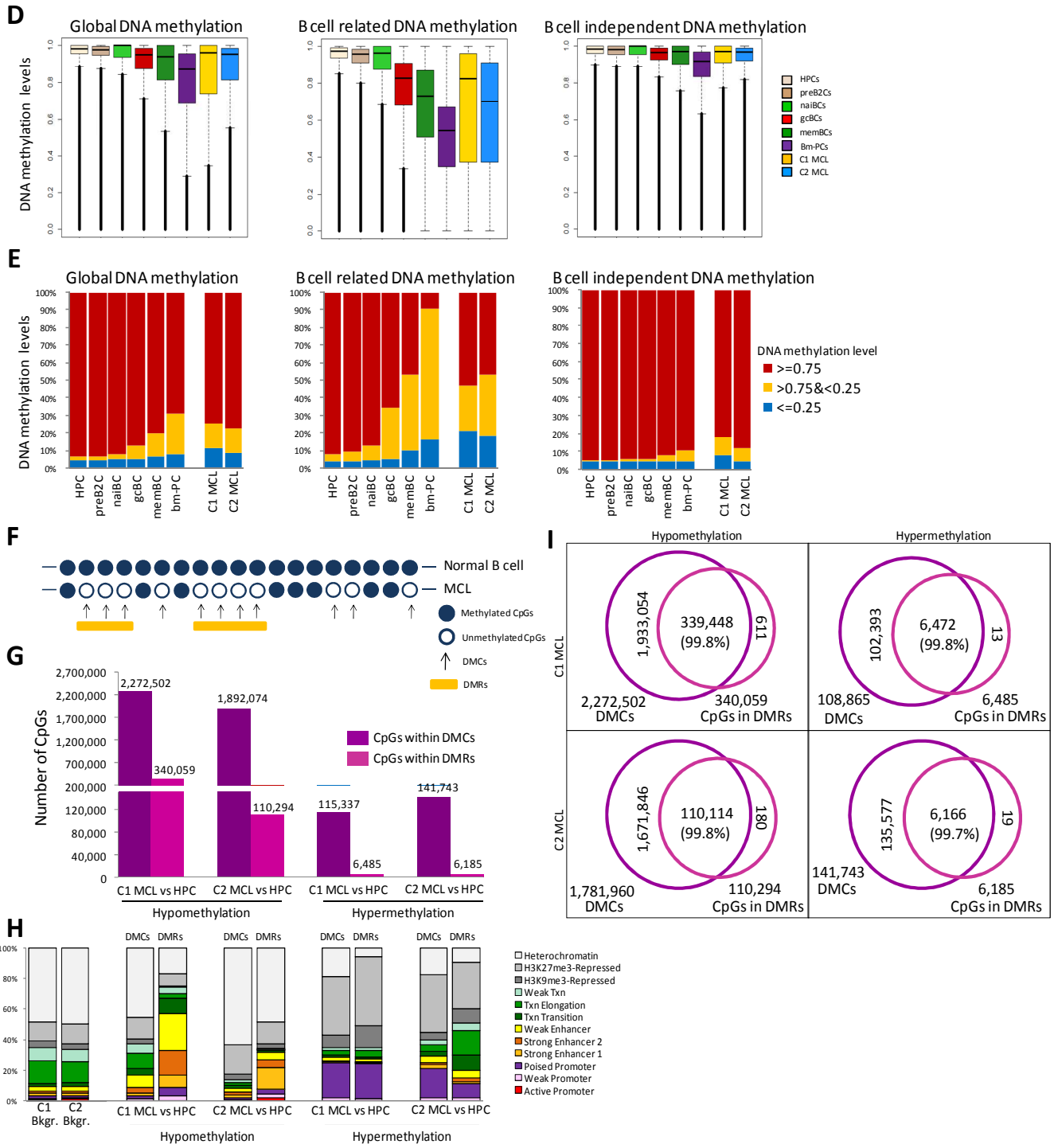


Figure S5, related to Figure 5. WGBS data.

(A) Identification of the two representative MCL cases used for WGBS in an unsupervised PCA of microarray methylation data. The two black arrows points to the two MCLs that were further analyzed by WGBS. **(B)** Heatmap of the differentially methylated CpGs between C1 and C2 as defined in Figure 2 that overlap with the CpGs mapped by WGBS in the two representative MCL primary cases. **(C)** Scatter plots showing the correlation between the DNA methylation values measure by 450k microarrays and WGBS for the two MCL primary cases. **(D)** Box plots summarizing the distribution of DNA methylation levels per sample for the 10.4 million CpGs with methylation estimates in all 8 samples. **(E)** Global distribution of DNA methylation in each sample showing the percentage of methylated (red), partially

methylated (yellow) and unmethylated (blue) CpGs. For both panels A and B, all CpGs are shown on the left, B cell-related CpGs in the middle and B cell-independent CpGs on the right. **(F)** Graphical representation of the strategies used to define differentially methylated CpGs (DMCs) and differentially methylated regions (DMRs). DMCs were defined at a single base resolution whereas DMRs where characterized by a minimum of 100bp and 3 CpGs in a row that showed the same tendency in DNA methylation. **(G)** Number of differentially methylated CpGs detected by DMCs and DMRs methods. **(H)** Comparison of the chromatin states of the differentially methylated CpGs detected by DMCs and DMRs methods. **(I)** Venn diagrams showing the overlap of CpGs detected by each of the strategies.

Table S3, related to Figure 5. Sequencing amounts in each of the two MCL samples analyzed by WGBS.

Sample name	Yield passing filter (Gb)	Mapping (%)	Unique mapping (%)	Mean coverage
C1 MCL	155.889	91.8	76.9	41.3
C2 MCL	208.840	91.0	72.3	53.8

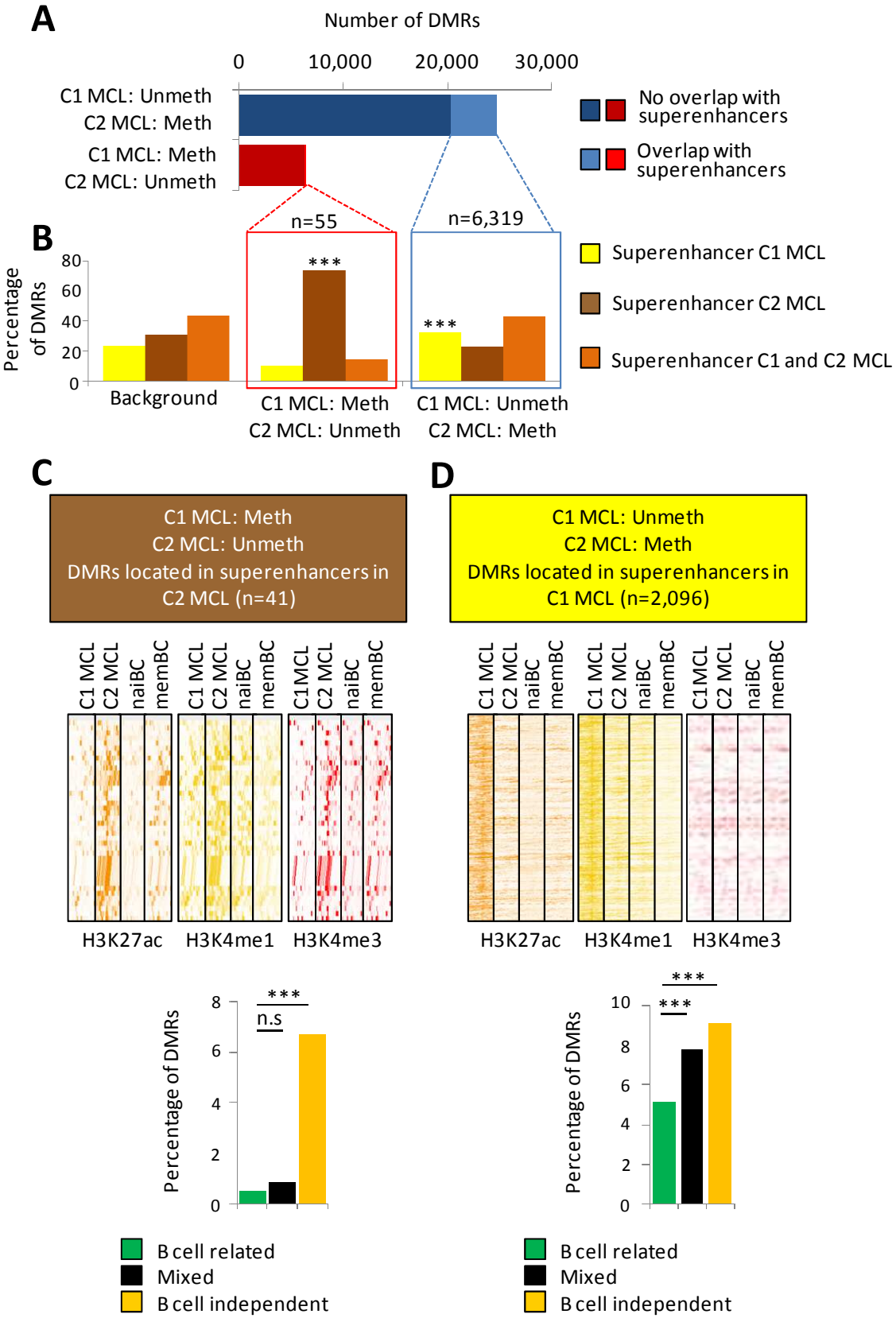


Figure S6, related to Figure 6. Analysis of super-enhancers.
(A) Number of DMRs between the C1 (SOX11-positive) and C2 (SOX11-negative) MCL cases and their overlap with super-enhancers in these MCL cases. **(B)** Distribution of the DMRs showing an overlap with super-enhancers in the C1 MCL case only, the C2 MCL case only or in both cases. The background represents all super-enhancers in the C1 and C2 MCL

cases, and shows which percentage is unique for these cases (yellow and darkbrown) and which percentage overlaps (lightbrown). ***Significantly higher than the background, P value<0.001, as determined by a Fisher's test. **(C-D)** Heatmaps showing the read density of H3K27ac, H3K4me1 and H3K4me3 chIP-seq in the C1 case, the C2 case, naive B cells (naiBC) and memory B cells (memBC) at selected DMRs (±25.000bp). Only the DMRs showing significant differences versus the background in panel B were used for these heatmaps. These are the unmethylated regions in the C2 case that overlap with super-enhancers in the C2 case only **(C)** and the unmethylated regions in the C1 case that overlap with super-enhancers in the C1 case **(D)**. In the lower part of these panels, the percentage of these respective DMRs within the B cell-related, mixed and B cell-independent DMRs is represented, showing a significantly higher (***$p < 0.001$) percentage in the mixed and B cell independent DMRs. n.s, non-significant.

Table S4, related to Figure 6. Differentially methylated regions that overlap with H3K27ac peaks in MCL primary cases. Provided as an Excel file.

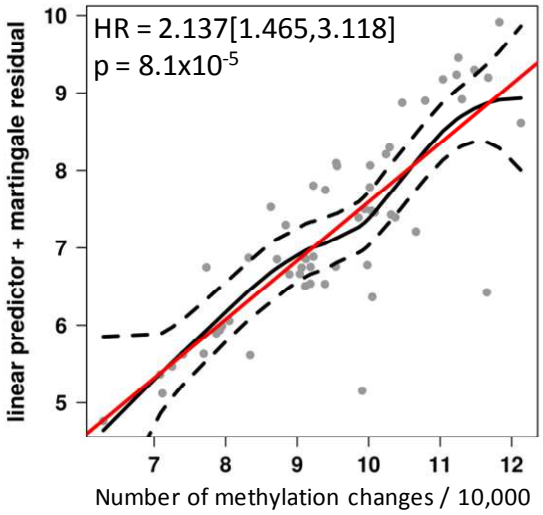
Table S5, related to Figure 6. Differentially methylated regions that overlap with super-enhancers in MCL primary cases. Provided as an Excel file.

Table S6, related to Figure 7. DNA methylation of CpGs located in the SOX11 distal enhancer region in MCL primary cases detected by bisulfite pyrosequencing.

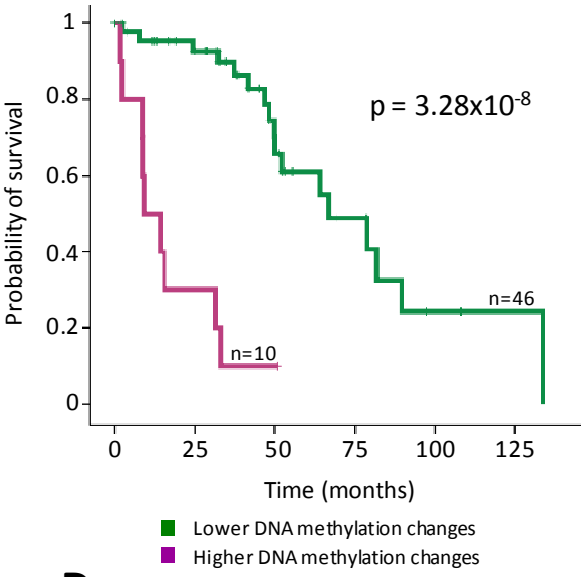
MCL Samples/ naiBC Samples	SOX11 status	chr2:6,477,577*	chr2:6,477,615*	chr2:6,484,702*	chr2:6,484,925*
M027	SOX11-negative	96	82	94	89
M039	SOX11-negative	78	44	12	11
M016	SOX11-negative	96	77	90	59
M163	SOX11-negative	93	61	82	ND
M197	SOX11-negative	97	84	59	91
M021	SOX11-negative	59	21	22	26
M218	SOX11-negative	55	34	ND	ND
M220	SOX11-negative	95	78	91	94
M228	SOX11-negative	91	75	79	80
M003	SOX11-negative	95	75	91	92
M029	SOX11-positive	3	8	21	14
M030	SOX11-positive	1	7	2	3
M035	SOX11-positive	30	20	17	21
M071	SOX11-positive	32	24	ND	ND
M078	SOX11-positive	7	10	10	10
M088	SOX11-positive	12	13	9	ND
M106	SOX11-positive	27	27	ND	ND
M108	SOX11-positive	ND	ND	7	10
M114	SOX11-positive	14	17	14	17
M198	SOX11-positive	21	19	18	20
M207	SOX11-positive	66	51	25	30
M212	SOX11-positive	ND	ND	ND	16
naiBC1	Non applicable	95	85	96	95
naiBC2	Non applicable	95	84	96	96
naiBC3	Non applicable	85	68	73	69
naiBC4	Non applicable	91	80	97	97

* DNA methylation levels are expressed as percentage

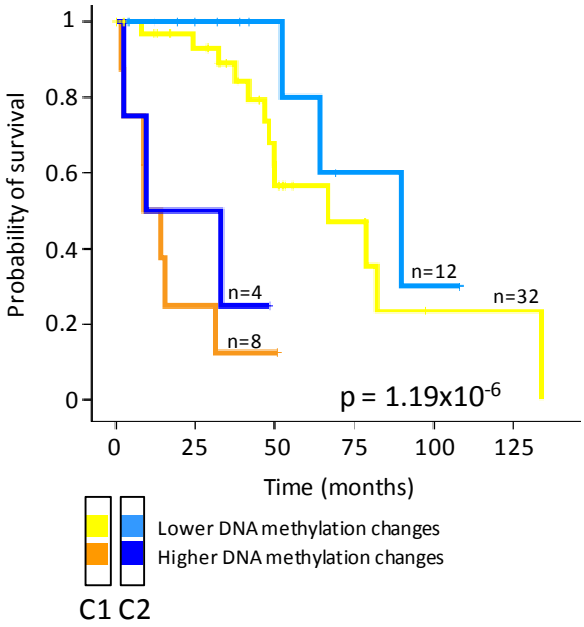
A



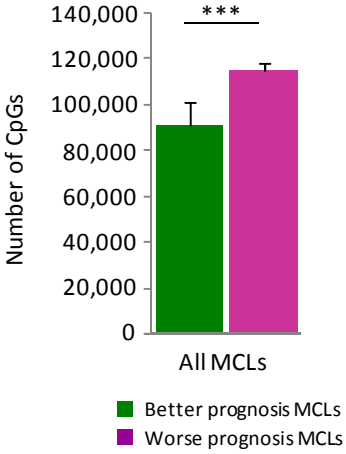
B



C

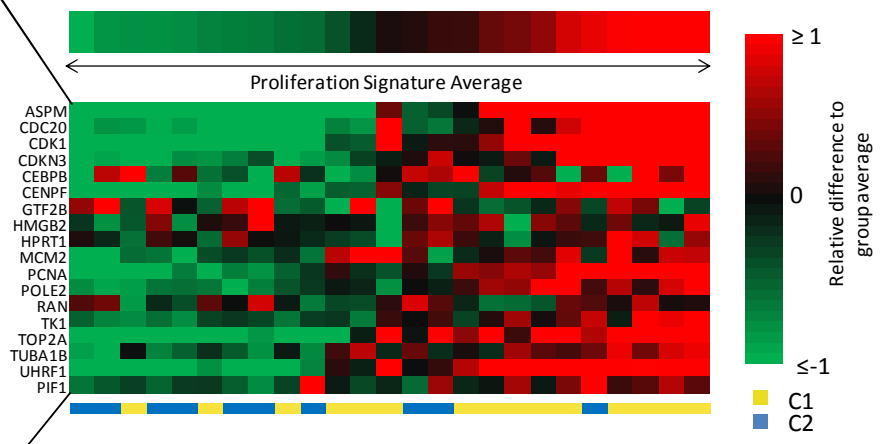


D



E

Rosenwald et al 2003	*HG U133 plus 2.0 array*	*Probeset HG U133 plus 2.0 array*
ASPM	ASPM	219918_s_at
p55CDC	CDC20	202870_s_at
CDC2	CDK1	203213_at
CIP2	CDKN3	209714_s_at
NF-IL6	CEBPB	212501_at
CENP-F	CENPF	207828_s_at
TFIIB	GTF2B	208066_s_at
HMG2	HMGB2	208808_s_at
HPRT	HPRT1	202854_at
MCM2	MCM2	202107_s_at
PCNA	PCNA	201202_at
DNApolE2	POLE2	205909_at
RAN	RAN	200749_at
LC34790	TK1	1554408_a_at
Topoisome-raseII alpha	TOP2A	201291_s_at
tubulin-alpha	TUBA1B	201090_x_at
UHRF1	UHRF1	225655_at
DNA helicase PIF1	PIF1	228252_at
FLJ10858	NEIL3	219502_at
LC26191	TDRD9	228285_at



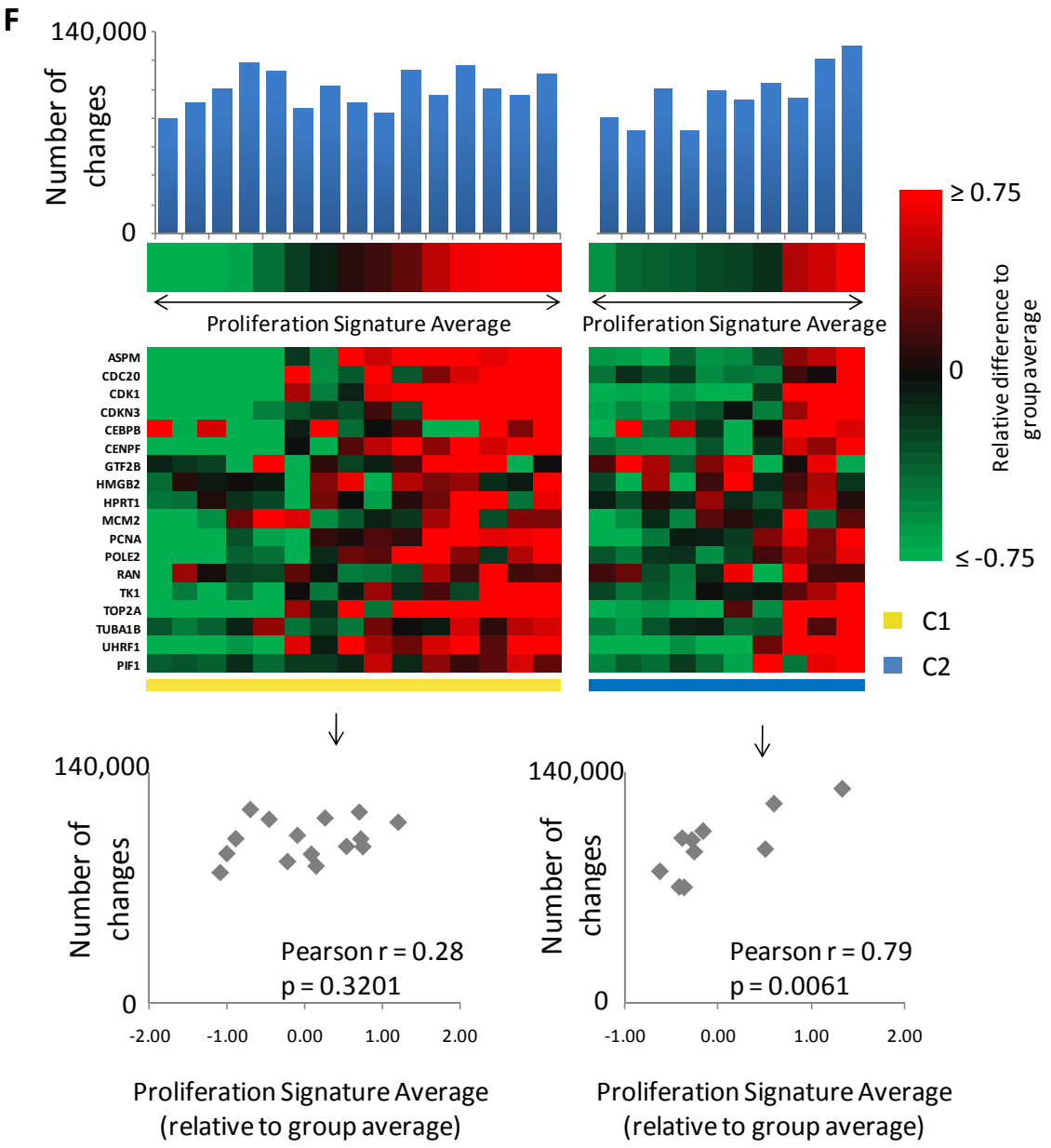


Figure S7, related to Figure 8. Prognosis and proliferation signature.
(A) Relationship between the number of epigenetic changes and overall survival in all MCLs through a linear predictor. Red line: perfect linear relationship; black line: local regression line; dash line: 95% confidence interval of local regression. **(B)** Kaplan-Meier plots of MCLs with lower vs. higher number of differentially methylated CpGs compared to HPCs. **(C)** Merged Kaplan-Meier plots with lower vs. higher number of differentially methylated CpGs compared to HPCs in C1 MCL and C2 MCL from main Figure 8C. Yellow and orange lines refer to C1 MCL cases whereas light and dark blue lines refer to C2 MCL cases. **(D)** Number of differentially methylated CpGs between the subgroups with different prognosis defined in panel B. Error bars show mean ± SD. ***p < 0.001 (t-test for independent samples). **(E)** Proliferation signature average of 25 MCL cases (columns) using the genes and probesets indicated in the last 2 columns of the left panel. NEIL3 and TDRD9 were excluded as they did not show expression in the MCL cases. **(F)** Proliferation signature of 15 (C1) and 10 (C2) MCL cases (heatmaps, using the same probesets as in panel E), the number of DNA methylation changes (upper panels, bar graphs) and the correlation between the proliferation signatures and the number of changes (lower 2 panels) in C1 (left) and C2 (right). P values were calculated using the Student's t-distribution.

Supplemental Experimental Procedures

MCL samples

The samples from MCL patients (n=82) were obtained from the Tumor Bank of the Hospital Clinic in Barcelona and the Institute of Human Genetics in Kiel. Clinical and biological characteristics of the patients are shown in Table S1. The enrolled patients gave written informed consent for scientific study according to the guidelines of the local Ethic Committees and the International Cancer Genome Consortium (ICGC) guidelines. From all the samples, genomic DNA was extracted using the QIAmp DNA kit (Qiagen) and assessed DNA quality by SYBR green staining on agarose gels quantified using the Nanodrop ND-1000 spectrophotometer (Thermo scientific).

MCL cell lines

The t(11;14)(q13;q32)-positive MCL cell lines Z-138, Granta-519 and JEKO-1 (SOX11- positive) and JVM-2 (SOX11-negative) were used as cellular models of MCLs. These cell lines were maintained in RPMI 1640 or DMEM, 10% FBS and Penicilin/Streptomicin at 37^{o}C with 5% CO_2. Cell lines tested negative for mycoplasma contamination, and their authenticity was confirmed by STR analysis using the CELL ID kit (Promega).

Normal B cells

Microarray-based and WGBS data from 10 sorted B cell subpopulations spanning the entire B cell differentiation, i.e., uncommitted hematopoietic progenitor cells, preBI cells, preBII cells, immature B cells, naive B cells from tonsils and peripheral blood, germinal center B cells, memory B cells from peripheral blood and plasma cells from tonsils and bone marrow (3-9 replicates of each), were used as controls in the present study (Kulis et al., 2015). WGBS data can be accessed through the European Genome-phenome Archive (EGA) under accessions EGAD00001001304 and EGAS00001000272. Microarray data are available under accession EGAS00001001196. Normal naive and memory B cells for 4C-seq were sorted from peripheral blood as previously described (Kulis et al., 2015).

Clinical and molecular characterization of MCL samples

IGHV-IGHD-IGHJ rearrangements and mutational status were analyzed using leader or consensus primers for the *IGHV* FR1 along with appropriate consensus primers,

as previously described (Navarro et al., 2012). Sequences with ≥98% identity to the germ line were considered unmutated. SOX11 expression was evaluated either by quantitative RT-PCR, gene expression profiling or immunohistochemistry and categorized as positive or negative as previously described (Navarro et al., 2012). Mutations in *TP53*, *WHSC1*, *BIRC3*, *NOTCH2*, *MEF2B* and *TLR2* genes were detected by whole-exome sequencing or by Sanger sequencing and analyzed with the Mutation Surveyor® (Softgenetics) (Bea et al., 2013).

Deconvolution and adjustment of DNA methylation estimates

In order to estimate the purity for all MCL samples we used a statistical approach formally presented in by Houseman and coworkers in 2012 (Houseman et al., 2012) and later adapted for 450k Illumina arrays by Jaffe and Irizarry in 2014 (Jaffe and Irizarry, 2014), made accessible through the function estimateCellCounts in the minfi package (version 1.18.2) available at Bioconductor. This function allows to determine the proportion of the main cell fractions of leukocytes in peripheral blood (B cells, $CD8^+$ T cells, $CD4^+$ T cells, NK cells, Monocytes and Granulocytes). As we report in our study, MCL has a highly disrupted DNA methylome, which in part affects CpGs whose methylation is modulated during B cell differentiation, and therefore, the function developed for normal peripheral blood samples may not be accurate in our experimental setting. We initially compared the B cell fraction estimated in silico and FACS data from 32 MCLs as gold standard, and observed an underestimation of B cell purity (data not shown). Therefore, we modified the published algorithm (Jaffe and Irizarry, 2014), and added the functionality to select a priori a set of CpGs for the deconvolution process and the capacity to remove unwanted CpGs, as well as the possibility to extract which set of CpGs are being used for the deconvolution. We proceeded selecting MCLs samples with high purity (≥99, based on FACS data) and comparing them to flow-sorted purified B cells (Reinius et al., 2012). We retained those CpGs showing similar methylation values (i.e. a difference ≤ 0.1 in methylation levels, n=184,547 CpGs). These CpGs represent those whose methylation levels globally remain stable in MCL lymphomagenesis and therefore are useful for the deconvolution process. We next provided those CpGs to the modified version of the aforementioned algorithm and obtained a more precise tumour B cell content estimate (Pearson coefficient between in silico estimation and FACS data of 0.95, $p < 0.001$) than with the initial approach (Jaffe and Irizarry, 2014) (Figure 1B). It is important to note that we improved the accuracy when predicting tumor B cell content

while we maintained the precision in discriminating the non-B cells of the microenvironment (Figure 1A). Afterwards, we assumed that the methylation estimates of MCL samples were the sum of the methylation estimates of the different cell types (Reinius et al., 2012) averaged by their calculated proportions:

$$Y = \sum_{1}^{n} \beta\theta = \beta_1\theta_1 + \beta_2\theta_2$$

where *Y* is the measured DNA methylation of the MCL samples, *n* is the total number of cell type fractions (6), β_1 is the in silico purified MCL DNA methylation estimates, θ_1 is the tumour B cell fraction, β_2 is the DNA methylation estimates of the non-tumour B cell and θ_2 is the proportions of non-tumor B cell, respectively. We proceeded with all downstream analysis of our study with this computationally purified methylation estimates (β_1) in all MCL samples (Figure 1D). Importantly, we verified that the large source of variability in DNA methylation due to cell compositions that was present with the initial methylation estimates, was completely removed after adjusting for MCL purities, and thus avoiding possible misleading associations coming from different purities in MCL samples (Figure S1).

DNA methylation analysis with 450K arrays

The EZ DNA Methylation Kit (Zymo Research) for bisulfite conversion was used to convert 500 ng genomic DNA per sample. Bisulfite-converted DNA was hybridized onto the HumanMethylation 450K BeadChip kit (Illumina) which covers 99% of RefSeq genes and 96% of CpG islands. The Infinium methylation assay was carried out as previously described (Bibikova et al., 2011; Bibikova et al., 2009).

Data from the 450k arrays were analyzed in R using the minfi package (version 1.18.2) (Aryee et al., 2014), available through the Bioconductor open source software. To exclude technical and biological biases that might produce false results in further analyses, we developed and optimized an analysis pipeline with several filters (i.e. discarding CpGs with low detection P values, sex-specific CpGs, CpGs showing individual-specific methylation and CpGs overlapping with SNPs). Taking into account the different performance of Infinium I and Infinium II assays we used the subset-quantile within array normalization (SWAN) (Maksimovic et al., 2012) that corrects for the technical differences between the Infinium I and II assay designs and produces a smoother overall beta value distribution.

Principal component analysis (PCA) and identification of MCL subgroups

PCA was performed using the prcomp function (Stats package, R software). Using linear discriminant analysis at each component individually we identified a cut-off value comparing the germinal center-inexperienced B cells versus germinal center-experienced B cells. We used this cut-off value to identify which MCL samples were more similar to germinal center-inexperienced B cells and which were more similar to germinal center-experienced B cells.

Whole-genome bisulfite sequencing (WGBS) and library construction

WGBS was performed on two MCL samples, one from each of the defined epigenetic groups (C1 MCL and C2 MCL). Genomic DNA (1–2 µg) was spiked with unmethylated λ DNA (5 ng of λ DNA per µg of genomic DNA) (Promega). The DNA was shared by sonication to 50–500 bp using a Covaris E220 and fragments of size 150-300 bp were selected using AMPure XP beads (Agencourt Bioscience Corp.). Genomic DNA libraries were constructed using the Illumina TruSeq Sample Preparation kit (Illumina Inc.) following the Illumina standard protocol: end repair was performed on the DNA fragments, an adenine was added to the 3' extremities of the fragments and Illumina TruSeq adapters were ligated at each extremity. After adaptor ligation, the DNA was treated with sodium bisulfite using the EpiTexy Bisulfite kit (Qiagen) following the manufacturer's instructions for formalin-fixed and paraffin-embedded (FFPE) tissue samples. Two rounds of bisulfite conversion were performed to ensure a conversion rate of over 99%. Enrichment for adaptor-ligated DNA was carried out through 7 PCR cycles using the PfuTurboCx Hotstart DNA polymerase (Stratagene). Library quality was monitored using the Agilent 2100 BioAnalyzer (Agilent), and the concentration of viable sequencing fragments (molecules carrying adaptors at both extremities) estimated using quantitative PCR with the library quantification kit from KAPA Biosystem. Paired-end DNA sequencing (2x100bp) was then performed using the Illumina Hi-Seq 2000. Amounts of sequence reads and the proportion of aligned reads are shown in Table S3.

Read mapping and calculation of cytosine methylation estimates

Read mapping was carried out using the GEM aligner (v1.242) (Marco-Sola et al., 2012) against a composite reference containing two copies of the human GRCh37 reference and two copies of the NCBI viral genome database (v35). For both the human and viral references, one copy had all C bases replaced by T and the other

had all G bases replaced by A. The names of the contigs in the combined reference FASTA file were modified by adding #C2T or #G2A to the end of the contig names depending on the conversion performed. Before mapping was performed the original sequence of each read was stored. The first read of each pair then had all C bases replaced by T, and the second read had all G bases replaced by A. Read mapping with GEM was performed allowing up to 4 mismatches per read from the reference. After read mapping the original sequence of each read was restored.

Estimation of cytosine levels was carried out on read pairs where both members of the read mapped to the same contig with consistent orientation, and there was no other such configuration at the same or less edit distance from the reference. After mapping, we restored the original read data in preparation for the inference of genotype and methylation status. We estimated genotype and DNA methylation status simultaneously using software developed at the Centro Nacional de Análisis Genómico (CNAG, Barcelona, Spain), taking into account the observed bases, base quality scores and the strand origin of each read pair. For each genome position, we produced estimates of the most likely genotype and the methylation proportion (for genotypes containing a C on either strand). A phred scaled likelihood ratio for the confidence in the genotype call was estimated for the called genotype at each position. For each sample, CpG sites were selected where both bases were called as homozygous CC followed by GG with a Phred score of at least 20, corresponding to an estimated genotype error level of <= 1%. Sites with >500x coverage depth were excluded to avoid centromeric/telomeric repetitive regions. A common set of called CpG sites for all analyzed samples was generated, and all subsequent analyses used this common set (n=11,384,077 CpGs).

Bisulfite pyrosequencing

Bisulfite pyrosequencing analysis was performed as previously described (Tost and Gut, 2007). Briefly, genomic DNA was bisulfite converted using EpiTectPlus Bisulfite Conversion Kit (Qiagen, Germany) according to manufacturer's specifications. Subsequent PCR amplification was performed using biotinylated primers. Pyrosequencing and data analysis were performed with the pyrosequencer analyzer PyroMark Q96 (Qiagen, Germany) according to manufacturer's instructions. The CpGs analysed and primer sequences used are detailed in the table below.

CpG coordinate (hg19)	Amplification primer (fw)	Amplification primer (rv)	Sequencing primer
chr2:6477577	/5Biosg/TTTATTGTTTTATAGTAAGGGTAGAG	CTAATCAAATACTCCCTAACC	AAAAATTCTAAATAATAACTCCTAC
chr2:6477615	/5Biosg/TTTATTGTTTTATAGTAAGGGTAGAG	CTAATCAAATACTCCCTAACC	AAAAATTCTAAATAATAACTCCTAC
chr2:6484702	GAGAAGTTGGTTTTAATGAGATTTAGTAGT	/5Biosg/CAAAAAAAAACCTTTAAAAAACAATACACC	GTTAATTTAAGTGGTTTTTGTTAT
chr2:6484925	ATTGTTGGAGATATGAGAAGTGT	/5Biosg/CCAAAACCTCATCTAAACCTACTTATTTC	ATTGTATTTTTTGAAGTTTTAATT

Differential DNA methylation analyses

To define regions with differential methylation between different MCL subgroups or between MCLs and normal B cell controls, using the 450K array data, we used (1) an absolute difference of mean DNA methylation levels of at least 0.25 between the compared groups and (2) a false discovery rate (FDR) of less than 0.05 using a Wilcoxon test for independent samples.

In the case of WGBS, we used two different strategies. On the one hand, differentially methylated CpGs (DMCs) were identified in a pair-wise comparison of each MCL sample versus HPCs, and the 2 MCL samples versus each other. Statistical significance difference in DNA methylation was estimated based in beta-binomial distribution using the "bdiff" algorithm (Raineri et al., 2014), and a DNA methylation difference of > 0.25. Annotation of CpGs was performed using the UCSC Table Browser GRCh37/hg19 version (Karolchik et al., 2004) and considering each feature related to a gene. On the other hand, differentially methylated regions (DMRs) were calculated using an Hidden Markov Model to segment the methylation values in the two samples under consideration. The Markov Model has 3 states corresponding to low, intermediate and high methylation; the transition probabilities are 0.9 for staying in the same state and 0.1 to change state. The emission probabilities of the HMM are the probabilities of obtaining the observed count of non-converted and converted reads assuming an underlying methylation value of less than 0.3 ("low" state), between 0.3 and 0.7 ("intermediate" state) and higher than 0.7 ("high" state). Stretches in the genome which correspond to the first samples being in "high" state and the second sample being in "low" state (or *vice versa*) are candidates for being DMRs. We further filtered these regions imposing that they contain more than 3 CpGs and that the difference in average methylation is larger than 0.25.

Both in the case of WGBS and 450K arrays, we classified differentially methylated sites into i) B cell-related CpGs (those whose DNA methylation level is modulated during the normal B cell differentiation), or ii) B cell-independent CpGs (those whose DNA methylation levels do not change during normal B cell differentiation). This classification was made based on recently published DNA methylation data from normal B cell subpopulations (Kulis et al., 2015).

In the case of WGBS, we further classified DMRs based on their composition of B cell-related and B cell-independent CpGs. We defined three classes of DMRs: i) B cell-related DMR (all CpGs within the DMR are B cell-related), ii) B cell-independent DMR (all CpGs in the DMR are B cell-independent) and iii) Mixed DMR (the DMR is formed by both B cell-related and B cell-independent CpGs).

Analysis of the DNA methylome of individual MCL samples

We studied the DNA methylation pattern of individual MCL in the context of the DNA methylome of B cell subpopulations spanning the entire differentiation program. This analytic strategy is independent from the maturation stage of the putative cell of origin of MCL and allows us to tackle individual epigenetic heterogeneity. We compared the DNA methylation profile of each individual MCL case with the mean DNA methylation per CpG of hematopoietic progenitor cells (HPCs) as fixed reference point. To classify a CpG as differentially methylated, we used a cut-off of an absolute difference of methylation values of at least 0.25. We also classified CpGs according to their degree of recurrence by overlapping differentially methylated CpGs from different individual cases.

Genomic and functional annotation of CpGs

Both CpGs analysed by WGBS and 450K array data were annotated using the UCSC Genome Browser database (hg19). For the location relative to a gene, we used the following categories: (i) TSS 1500 (from 201 to 1,500 bp upstream of the transcriptional start site (TSS)), (ii) TSS 200 (from 1 to 200 bp upstream of the TSS), (iii) 5′ UTR, (iv) first exon, (v) gene body (from the first intron to the last exon), (vi) 3′ UTR and (vii) intergenic regions. Owing to the presence of alternative transcription start sites and regions containing more than one gene, some of the CpGs were assigned multiple annotations. For the location relative to a CpG island (CGI), we used the following groups: (i) within CGI, (ii) in CGI shore (0–2 kb from the CGI edge), (iii) in CGI shelf (>2 kb to 4 kb from the CGI edge) and (iv) outside CGI.

We also annotated all CpG probes according to chromatin activity by mapping chromatin states in normal naive (n=1) and memory B cells (n=1), as well as in 2 representative MCL cases (for which also the WGBS and 4C-sequencing was performed). To that end we used the chromHMM software (Ernst and Kellis, 2012) using ChIP-seq data of 6 histone marks (H3K4me1, H3K4me3, H3K27ac, H3K36me3, H3K27me3, H3K9me3) as input. The following regions were considered, adapted from a previously published segmentation (Ernst et al., 2011): Active promoter ($H3K4me3^{+}$, $H3K27ac^{+}$), Weak Promoter ($H3K4me1^{+}$, $H3K4me3^{+}$), Poised Promoter ($H3K4me1^{+}$, $H3K4me3^{+}$, $H3K27me3^{+}$), Strong Enhancer 1 ($H3K4me1^{+}$, $H3K4me3^{+}$, $H3K27ac^{+}$), Strong Enhancer 2 ($H3K4me1^{+}$, $H3K27ac^{+}$), Weak Enhancer ($H3K4me1^{+}$), Transcription Transition ($H3K36me3^{+}$, $H3K4me1^{+}$, $H3K27ac^{+/-}$), Transcription Elongation ($H3K36me3^{+}$), Weak Transcription ($H3K36me3^{+/-}$), H3K9me3 Repressed ($H3K9me3^{+}$), H3K27me3 Repressed ($H3K27me3^{+}$), Heterochromatin Low signal (none of the 6 marks).

Correlation of DNA methylation with somatic hypermutation levels

For C1 and C2 MCLs separately, DNA methylation levels of each CpG and the level of somatic hypermutation (SHM) were correlated using spearman correlation tests using the rcorr function of the Hmisc R package. SHM levels ranged from 0 (no SHM, i.e. 100% germline identity) to 12.32 (high level of SHM, 87.68% identity with germline). Significant correlations between CpG methylation and SHM were detected as presenting an absolute correlation coefficient > 0.4 and an FDR-value < 0.05. The R statistical package version 3.1.1 was used for these analyses.

ChIP-seq experiments and analysis

ChIP-Seq of H3K4me1, H3K4me3, H3K27ac, H3K36me3, H3K27me3, and H3K9me3 were performed using standard protocols generated within the Blueprint Consortium. Protocol details can be found at: http://www.blueprint-epigenome.eu/index.cfm?p=7BF8A4B6-F4FE-861A-2AD57A08D63D0B58. Catalog numbers of antibodies (Diagenode) used are H3K27ac: C15410196/pAb-196-050 (LOT: A1723-0041D), H3K4me1: C15410194/pAb-194-050 (LOT: A1863-001P), H3K4me3: C15410003-50/pAb-003-050 (LOT: A5051-001P), H3K36me3: C15410192/(pAb-192-050 (LOT: A1847-001P), H3K9me3: C15410193/pAb-193-050 (LOT: A1671-001P), H3K27me3: C15410195/pAb-195-050 (LOT: A1811-001P). Regions significantly enriched for H3K27ac were determined using MACS2, using

default setting (Zhang et al., 2008). The read density of H3K27ac, H3K4me1 and H3K4me3 were visualized with seqMiner (Ye et al., 2014).

4C-sequencing experiments and analysis

4C templates were prepared as previously described (Simonis et al., 2007; van de Werken et al., 2012). Briefly, $1x10^7$ cells (MCL primary cases, GRANTA-519, JEKO-1, Z-138 and JVM-2) or $4x10^6$ cells (naive and memory B cells) were crosslinked with 2% (MCL primary cases, GRANTA-519, JEKO-1, Z-138 and JVM-2) or 1% (naive and memory B cells) formaldehyde, chromatin was digested with a first restriction enzyme (NlaIII) followed by ligation. Next, chromatin was decrosslinked and DNA was digested with a second restriction enzyme (DpnII) and ligated. PCR amplification of viewpoint regions and its ligated fragments was performed using the 4C templates. The alternative protocol for naive and memory B cells was used due to the availability of lower cell numbers of these samples. It was tested using Z-138 ($2x10^6$ cells) and we observed that with the new protocol the interaction between the *SOX11* promoter and its distant enhancer was well detectable in Z-138. Samples were sequenced with the MiSeq instrument (Illumina) using 50bp single-reads, adding 5% PhiX. Restriction enzymes and primers used are indicated in the table below. 4C-seq analysis was performed using the 4C-seq pipeline 4cseqpipe (http://compgenomics.weizmann.ac.il/tanay/?page_id=367). The R statistical package version 3.1.1 was used for these analyses.

Region	Viewpoint Fragment (hg19)	RE1	RE2	RE1_primer	RE2_primer
SOX11 locus	chr2:5834180-5835254	NlaIII	DpnII	CCACCAAAATTTTCATCATG	TCTTCTATGCATCCGATTCT
SOX11 enhancer	chr2:6492207-6492728	NlaIII	DpnII	TCAGACTGACTTTCCTCATG	TCTTCGTGTTTAAGATTCCC

RE1 = first restriction enzyme, RE2 = second restriction enzyme

Analysis of super-enhancers

Super-enhancers in the two representative MCL cases were detected using the ROSE software (Loven et al., 2013; Whyte et al., 2013). Thereby, the H3K27ac peaks, as determined by MACS2 were used as input, as well as the mapped reads of the H3K27ac ChIP-seq experiments and their respective input controls to correct for background levels. The H3K27ac peaks present at transcription start sites were removed from this analysis, by turning on the option –t in the ROSE software.

Gene ontology analysis

The KEGGprofile package (Zhao and Shyr, 2015) available through Bioconductor was used to determine the enrichment of individual ontology terms in the different methylation CpGs between C1 and C2 MCLs as compared to all the genes analyzed in the 450K array. Significant terms were considered when adjusted $p < 0.05$.

Analysis of the proliferation signature

Gene expression data of 25 MCL patients for which DNA methylation data was available were mined (Navarro et al., 2012). The expression levels of 18 genes were used to calculate the proliferation signature (Rosenwald et al., 2003), see also Figure S8. To that end, the mean of the rma normalized expression values of these 18 genes was calculated per sample. To calculate the relative difference of the proliferation signature compared to the group average, the average proliferation signature value of the group was subtracted from the proliferation signature per case. This average proliferation signature was then correlated with the number of DNA methylation changes in cases from C1 and C2 MCLs.

Statistical analyses

The relationships between MCL subgroups and clinical and biological variables of patients described in Figure 2 was evaluated using the Fisher's exact test in the case of qualitative variables and the t-test for independent samples in the case of quantitative variables with normal distribution (a corrected p value was used if the two groups had unequal variances). Statistical significance was defined as $p < 0.05$. The sample size in each of the comparisons varied depending on the available data.

The relationship between the number of epigenetic changes and overall survival was examined by plotting the number of changes (x-axis) vs. the linear predictor plus martingale residuals (y-axis). The local regression line suggests a linear relationship between the number of changes and the log hazard.

To detect MCL groups with different clinical behavior based on their DNA methylation changes, we used the "maxstat" package from R software. This analysis allows us to detect the most suitable threshold to separate MCL groups with differences in overall survival. Overall survival Kaplan-Meier plots and long-rank tests were performed with the IBM-SPSS Statistics version 20.

A multivariate analysis was performed to determined whether the epigenetic burden is an independent predictor of prognosis, using the coxph function (Survival package, R software). The multivariate Cox Regression model was generated using as input variables: age (quantitative), morphology (classical/blastoid), *IGHV* somatic hypermutation (quantitative), presence of mutations (positive/negative, based on the analysis of the 6 most frequent drivers), SOX11 expression (positive/negative) and number of DNA methylation changes (quantitative).

Supplemental References

Aryee, M. J., Jaffe, A. E., Corrada-Bravo, H., Ladd-Acosta, C., Feinberg, A. P., Hansen, K. D., and Irizarry, R. A. (2014). Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics *30*, 1363-1369.

Bibikova, M., Le, J., Barnes, B., Saedinia-Melnyk, S., Zhou, L., Shen, R., and Gunderson, K. L. (2009). Genome-wide DNA methylation profiling using Infinium(R) assay. Epigenomics *1*, 177-200.

Ernst, J., and Kellis, M. (2012). ChromHMM: automating chromatin-state discovery and characterization. Nat Methods *9*, 215-216.

Ernst, J., Kheradpour, P., Mikkelsen, T. S., Shoresh, N., Ward, L. D., Epstein, C. B., Zhang, X., Wang, L., Issner, R., Coyne, M.*, et al.* (2011). Mapping and analysis of chromatin state dynamics in nine human cell types. Nature *473*, 43-49.

Karolchik, D., Hinrichs, A. S., Furey, T. S., Roskin, K. M., Sugnet, C. W., Haussler, D., and Kent, W. J. (2004). The UCSC Table Browser data retrieval tool. Nucleic Acids Res *32*, D493-496.

Loven, J., Hoke, H. A., Lin, C. Y., Lau, A., Orlando, D. A., Vakoc, C. R., Bradner, J. E., Lee, T. I., and Young, R. A. (2013). Selective inhibition of tumor oncogenes by disruption of super-enhancers. Cell *153*, 320-334.

Maksimovic, J., Gordon, L., and Oshlack, A. (2012). SWAN: Subset-quantile within array normalization for illumina infinium HumanMethylation450 BeadChips. Genome Biol *13*, R44.

Marco-Sola, S., Sammeth, M., Guigo, R., and Ribeca, P. (2012). The GEM mapper: fast, accurate and versatile alignment by filtration. Nat Methods *9*, 1185-1188.

Raineri, E., Dabad, M., and Heath, S. (2014). A note on exact differences between beta distributions in genomic (Methylation) studies. PLoS One *9*, e97349.

Rosenwald, A., Wright, G., Wiestner, A., Chan, W. C., Connors, J. M., Campo, E., Gascoyne, R. D., Grogan, T. M., Muller-Hermelink, H. K., Smeland, E. B.*, et al.* (2003). The proliferation gene expression signature is a quantitative integrator of oncogenic events that predicts survival in mantle cell lymphoma. Cancer Cell *3*, 185-197.

Tost, J., and Gut, I. G. (2007). DNA methylation analysis by pyrosequencing. Nat Protoc *2*, 2265-2275.

Whyte, W. A., Orlando, D. A., Hnisz, D., Abraham, B. J., Lin, C. Y., Kagey, M. H., Rahl, P. B., Lee, T. I., and Young, R. A. (2013). Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell *153*, 307-319.

Ye, T., Ravens, S., Krebs, A. R., and Tora, L. (2014). Interpreting and visualizing ChIP-seq data with the seqMINER software. Methods Mol Biol *1150*, 141-152.

Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., Nusbaum, C., Myers, R. M., Brown, M., Li, W., and Liu, X. S. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol *9*, R137.

Zhao, S., and Shyr, Y. (2015). KEGGprofile: An annotation and visualization package for multi-types and multi-groups expression data in KEGG pathway.